Best practices for releasing code

For all experiments you report, check if you released:

☐ Code for the training pipeline used to evaluate the final architectures
☐ Code for the search space
☐ The hyperparameters used for the final evaluation pipeline, as well as random seeds
☐ Code for your NAS method
☐ Hyperparameters for your NAS method, as well as random seeds

Note that the easiest way to satisfy the first three of these is to use existing NAS benchmarks, rather than changing them or introducing new ones.

Best practices for comparing NAS methods

☐ For all NAS methods you compare, did you use exactly the same NAS benchmark, including the same dataset (with the same training-test split), search space and code for training the architectures and hyperparameters for that code?
☐ Did you control for confounding factors (different hardware, versions of DL libraries, different runtimes for the different methods)?
☐ Did you run ablation studies?
☐ Did you use the same evaluation protocol for the methods being compared?
☐ Did you compare performance over time?
☐ Did you compare to random search?
☐ Did you perform multiple runs of your experiments and report seeds?
☐ Did you use tabular or surrogate benchmarks for in-depth evaluations?

Best practices for reporting important details

☐ Did you report how you tuned hyperparameters, and what time and resources this required?
☐ Did you report the time for the entire end-to-end NAS method (rather than, e.g., only for the search phase)?
☐ Did you report all the details of your experimental setup?

For details on these best practices, please see our JMLR paper "Best Practices for Scientific Research on Neural Architecture Search“, [https://www.jmlr.org/papers/v21/20-056.html](https://www.jmlr.org/papers/v21/20-056.html)